UDC 62-50

A NOVEL APPROACH TO REMOTE SENSING OF VEGETATION

P.I. BIDYUK, V.I. LITVINENKO, S.O. PONOMARENKO

The problem of remote estimation of chlorophyll content in vegetation is considered. Some reflectance spectra have been recorded for winter wheat leaves with various level of chlorophyll concentration. To reduce the level of noise in the measurement data produced by measuring system and possible influence of soil surface smoothing procedure proposed by Savitzky and Golay was applied. The 1st derivative of reflectance spectra curves had been computed and analyzed with respect to correlation with pigment content. To compute an estimate of chlorophyll content multiple regression as well as neural net approach have been applied and both proved to be successful.

Introduction. Chlorophyll content in plant leaves is an important characteristic showing on the plant status. For example, pigment level correlates with power of photosynthetic apparatus and may be used for yield prognosis. Influence of various stresses on photosynthetic apparatus first of all induces changes in chlorophyll content. This phenomenon allows to develop effective approaches to fast revealing stresses in phytocenoses of various types, using remote estimates of chlorophyll content and its dynamics. It is well known that reflectance spectra of plant leaves in an optical range are the most informative with respect to chlorophyll content due to the fact that they are correlated with specific spectrum of the pigment. Moreover, these spectra may serve as reliable information for remote measurements.

Therefore, an idea to apply this characteristic to monitoring of agrocenosis and other types of plant canopies seems to be quite acceptable. There are numerous relations based on regression equations used to calculate chlorophyll concentration in leaves by making use of reflectance spectral coefficients or their combinations [1–6]. Practically all of them are effective enough when chlorophyll concentration is measured under laboratory conditions. When estimation of chlorophyll concentration is performed for «soil-vegetation» system then estimation results may be distorted. The main cause is in contribution of soil reflectance. There are several approaches to minimize this type of noise by using complex vegetation indices including components for correction of soil interference [7–9]. Their application demands measurements of additional parameters of the system, such as leaf index and spectral coefficients of soil reflectance. This approach makes the procedures low effective and moreover, the problem is not solved completely because of variability of soil reflectance depending on its moisture. Therefore, it is very important to develop effective methods acceptable for chlorophyll measurement in the «soil-vegetation» system.

Application of high resolution spectroscopy to remote sensing of vegetation [10] displayed new characteristics correlating with chlorophyll content. It appeared that a position of the so called *red edge* in spectral curves of reflectance spectra and the shape of this spectral region were dependent on pigment concentration. Various characteristics of red edge (680–800 nanometers) have been used.

Position of maximum in the 1st derivative plot of reflectance spectral curve [11] and position of red edge were found to be in regression relation with chlorophyll content [12–14]. It was also revealed that there exists correlation between the ratio of two maxima in the 1-st derivative plot and chlorophyll [15–16]. Some resistance of this characteristic to soil contribution has been shown as well [17]. A shortage of this approach is in its sensitivity to instrumental noises, which may considerably distort the results influencing, for example, intensity of maxima and their position. Besides, one of the two maxima is very often difficult to distinguish what creates extra computing burden.

Thus, to develop a device for remote measurements of chlorophyll content in vegetation using the 1st derivative characteristics it is necessary to solve the following problems: 1) to reduce the level of noise in the measurement data; 2) to propose appropriate variables and computing procedures for estimating chlorophyll content that would exhibit acceptable quality.

The methodology of spectral and biochemical measurements. Leaves of two varieties of winter wheat were studied during the experiments, and appropriate measurements had been collected. Variations of chlorophyll content were caused by differences in mineral nutrition and plant vegetation age. Reflectance spectra were measured with spectrophotometer SF-10 (USSR) equipped with integrating sphere [17]. The recording system allowed to obtain the spectra in digital form. The spectra were recorded with the increment of 1 *nm* in the range of 680–750 *nm*. The pieces of the wheat leaves used for recording of reflectance spectra, and the data were used for chlorophyll estimation by chemical method proposed by Arnon [18].

Preliminary processing of experimental data. Preliminary data processing is based on application of smoothing procedure proposed in [19, 20]. To perform data smoothing we used several moving windows of different size: 5, 7 and 9 point windows. After data smoothing the 1st derivative was computed using 2-d order polynomial estimated by making use of LS procedure. For this purpose another moving window was applied to smoothed data. Next, the resulting first derivative curve was used to determine estimates of chlorophyll content.

Estimation of chlorophyll content in vegetation. To determine estimates of chlorophyll content in vegetation the following approaches were applied: regression equations of various structure (order) and neural nets.

Preliminary smoothing of measurement data was performed with 7 and 9-point windows that revealed their acceptable effectiveness. After smoothing the data we computed the 1st derivative of reflectance spectra recorded for winter wheat leaves with various chlorophyll concentration. Some examples of the 1st derivative plots are shown in figure. Visual estimation of the plots shows that the ripples are insignificant. In this study we did not consider the two maxima approach reported earlier [16, 17]:

$$y = k_0 + k_1 \frac{x_{max}^{(2)}}{x_{max}^{(1)}}, \qquad (1)$$

where $x_{max}^{(1)}$, $x_{max}^{(2)}$ are two maxima of the first derivative of reflectance spectra. Though ratio of the maxima intensity is dependent on chlorophyll content in leaves. To determine coefficients k_0 and k_1 two reflectance spectra are required.

The approach had been studied in [16, 17], and proved to be useful to some extent. The drawback of this approach is that it uses only two points of the first derivative plot what creates a problem with reaching desirable quality of estimates of chlorophyll content in vegetation. Also, computing experiments with locating «lower» maximum showed that it was not always possible to find it. That is why we propose to use multiple regression equation considered below.
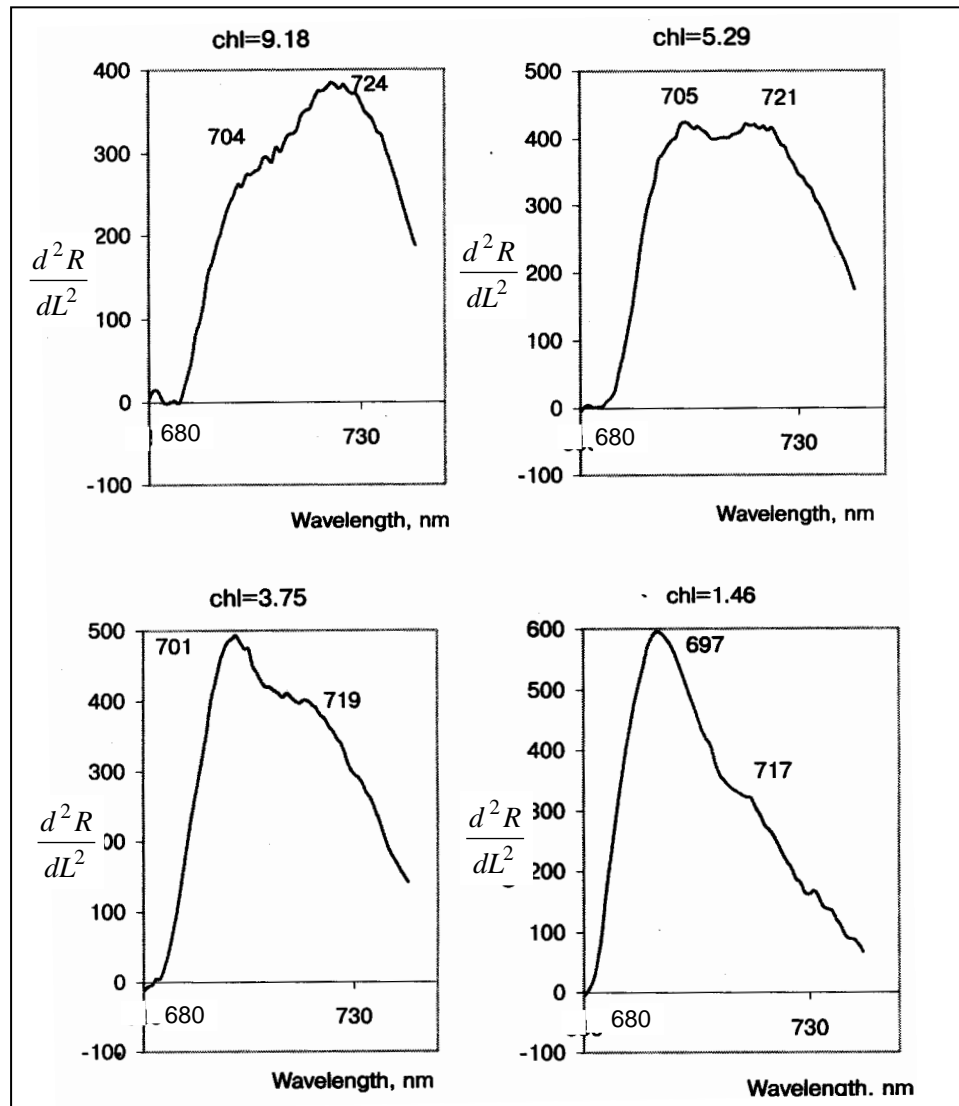
Plots of first derivatives of reflectance spectra for winter wheat leaves versus wavelength, *nm*

Regression equation. The regression equation proposed has the following form

$$z(l) = a_0 + a_1 x_1(l) + a_2 x_2(l) + ... + a_m x_m(l) + \varepsilon(l), \qquad (2)$$

where $x_i, i = 1,...,m$ are values of the first derivative of smoothed reflectance spectrum taken with appropriate interval that could be varied depending on the

regression order and number of experiments available; l is a number of an experiment, with $l = 1, 2, ..., N$; ε is equation error that is practically inevitable in estimating regression. Here one experiment is referred to one session of spectrum measurement using spectrometer, and to determining one estimate of chlorophyll concentration by chemical test. Values of the first derivative were taken in the interval $680-740$ *nm*. Thus, theoretically order of the regression could be expanded to about 60. However, in practice order of linear regression could be reduced to the value of 10–15. Simplicity of the equation allows to use it even in real time with signal processor implementation.

Number of independent variables in equation (2) is much higher than in equation (1) what makes it logically more suitable for description of the first derivative of measurement data. The larger is the number of independent variables the higher should be degree of adequacy of equation (2), but there are computational limits to expanding the right-hand side, and shortage of an interval between samples of spectra may result in linearly dependent vectors of measurement matrix and, consequently, in the so called co-linearity problem. To find estimates of coefficients a_i, $i = 0, 1, ..., m$ we need at least $m + 1$ measurements of reflected spectra and respective estimates of chlorophyll content.

Some results of applying equation (2) to estimation of chlorophyll content using experimental reflectance spectra are given in Tabl. 1.

T a b l e   1. Results of chlorophyll content estimation using multiple regression.

№ of exp.	Regres. coeff.	Chlorof. estimate	Error of estimate x 1E-08	Regres. coeff.	Chlorof. estimate	Error of estimate x 1E-06	Regres. coeff.	Chlorof. estimate	Error of Estimate x 1E-06
	1	2	3	4	5	6	7	8	9
1	–3.346	4.76	0.4912	–35.935	4.76	0.1134	–21.779	4.76	0.0498
2	–2.756	5.67	0.5575	30.508	5.67	0.1093	0.952	5.67	0.0168
3	9.865	3.07	0.3703	–14.511	3.07	0.1057	–10.075	3.07	0.1829
4	–12.996	3.58	0.5072	–32.932	3.58	0.0867	34.203	3.58	0.1630
5	19.376	6.82	0.5391	5.490	6.82	0.1154	–42.364	6.82	0.0542
6	–28.401	6.57	0.5695	56.401	6.57	0.1372	58.494	6.57	0.0178
7	30.747	8.02	0.4117	–127.459	8.02	0.1091	–79.216	8.02	0.0965
8	–28.317	3.75	0.5485	215.172	3.75	0.0913	69.327	3.75	0.1008
9	21.174	7.65	0.5663	–134.240	7.65	0.1378	–32.189	7.65	0.1013
10	–29.104	3.25	0.3816	57.537	3.25	0.0719	–46.969	3.25	0.2364
11	26.135	2.43	0.4817	–159.463	2.43	0.1160	61.573	2.43	0.0170
12	–4.337	5.04	0.5806	122.207	5.04	0.1079	–9.275	5.04	0.0958

According to Tabl. 1, estimates of chlorophyll, computed via equation (2), exhibit rather high quality that is acceptable for field studies of vegetation state. Columns 1, 4, and 7 contain regression coefficients computed for different values of samples that belong to the range of wavelength between 680–740 *nm*. Substantial difference between values of the coefficients indicates that the values of the first derivative, used for computing regression, should be carefully selected. Col-

umns 2, 5, and 8 show estimates of chlorophyll content computed with regression estimated via the first derivatives of reflectance spectra. The estimates have the same values due to the fact that the estimate errors (columns 3, 6, and 9) are very low for each of the three regressions tested in this study. The smallest values of estimate errors are presented in column 3 for the first regression what indicates that left edge of the first derivative curve is more informative than the right one.

Thus, this method seems to be promising for the field estimation of chlorophyll concentration in vegetation. Further study of the regression approach to estimation is required in the following directions: (1) vary the number of independent variables; (2) vary the range of wavelength used for estimation; (3) try various structures of regression including nonlinear one.

Neural net. As an alternative way to regression model we have examined possibility of application of a neural net to determining estimates of chlorophyll content. In many cases neural nets exhibit robustness with respect to noise of various origin, and provide acceptable estimates when regression works onsatisfactorily. A multilayer neural net approach was applied. It includes one input layer; two, three or more hidden layers, and one output layer. The input layer is a distributor of input sequence $\{1,...,j,...,J\}$ for n_1 neurons of the first hidden layer. Each neuron has its weighting coefficient $K_{j,n1}$ which form a matrix of weighting coefficients. The products formed by each neuron are summed and applied to the input of non-linear sigmoid type element. Outputs of the sigmoid elements are applied to inputs of the second hidden layer of the network and so on. The output layer of the net consists of a number of neurons, which corresponds to the number of classes each of which includes definite selected values of chlorophyll estimates. The large is the number of classes the lower is error of estimate of chlorophyll content. Number of neurons in each layer, number of layers, coefficients $K_{j,n1}$, functions of non-linear elements are adjusted in the process of constructing specific net. There are several approaches to determine minimum and maximum number of neurons for a net. To estimate a number of neurons in hidden layers we used the expression for estimating necessary number of synaptic weights [21, 22]:

$$\frac{mN}{1+\log_2 N} \le L_w \le m\left(\frac{N}{m}+1\right)(n+m+1)+m,\qquad(3)$$

where n is a dimension of input signal; m is dimension of output signal; N number of elements of training sample. Then necessary number of neurons in hidden layers for a two later net is:

$$L=L_w(n+m).\qquad(4)$$

Other parameters are determined during the training process when the net is fed with representative spectral data and estimates of chlorophyll content produced by chemical analysis. To determine the coefficient values and parameters of non-linear functions were used the training procedures given in Tabl. 2.

As an effectiveness indicator of the neural net functioning were selected statistical characteristics of net response to training and control sequences. The ap-

proach, based on a neural net, does not require a priory knowledge of decision rules. Though, depending on a net structure it could be constructed robust with respect to disturbances (noise) of various nature. It is also important to stress that they do not require estimating of a model structure what is usually done when regression approach is applied.

T a b l e 2. The training algorithms used

№	Type of training algorithm	Matlab function
1	BFGS quasi-Newton backpropagation	Trainbfg
2	Powell-Beale conjugate gradient backpropagation	Traincgb
3	Fletcher-Powell conjugate gradient backpropagation	Traincgf
4	Polak-Ribiere conjugate gradient backpropagation	Traincgp
5	Gradient descent backpropagation	Traingd
6	Gradient descent with adaptive of learning backpropagation	Traingda
7	Gradient descent with momentum backpropagation	Traingdm
8	Levenberg-Marcquardt backpropagation	Trainlm
9	One step secant backpropagation	Trainoss
10	Scaled conjugate gradients backpropagation	Trainscg

After training quality of operation of each of the networks studied was estimated by making use of mean squared error (MSE) computed for p samples of data:

$$MSE = \sqrt{\frac{1}{p}\sum_{i=1}^{p}(d_i - \widehat{y}_i)^2} \,,$$

where d_i is desirable value of output, and $\widehat{y}_i$ is a actual estimate found for i th sample. MSE was computed separately for a sample with known, MSE_c, and unknown (test sample), MSE_t, concentrations of chlorophyll. The results of estimation and number of training cycles are given in Tabl. 3. It follows from columns 3 and 4 that MSE_c and MSE_t are close to each other.

T a b l e 3. Results of estimation and neural nets training

№	Training function	MSE_t	MSE_c	№ of training cycles
1	Trainbfg	0.201	0.198	3120
2	Traincgb	1.263	1.351	6320
3	Traincgf	0.351	0.345	4550
4	Traincgp	1.725	1.801	5150
5	Traingd	2.101	2.214	21000
6	Traingda	2.478	2.541	7014
7	Traindm	3.001	3.124	19950
8	Trainlm	1.145	1.102	6651
9	Trainoss	1.996	1.897	3125
10	Trainscg	0.471	0.456	4100

Application of neural nets shows that generally they provide different results for different types of training algorithms. The best results in our experiments were achieved with application of BFGS quasi-Newton backpropagation and Fletcher-Powell conjugate gradient backpropagation optimization training procedure with MSE (Mean Squared Error) of estimate 0.201, and 0.471 for scaled conjugate gradient algorithm. In some cases (#5, and #7) number of iterative training cycles is very high what proves again that the problem of training time reduction is also present in this study. Also it is clear that application of neural nets still require further research with respect to refining net structure aiming to improve the estimates.

Conclusions. Application of multiple regression to estimating concentration of chlorophyll in vegetation led to positive results. Multiple linear regression was quite effective, and increasing the number of independent variables improved quality of estimates substantially. In a case of eleven independent variables the estimate errors turned out to be very small, what proves the possibility of using the equation in a field device for preliminary estimation of vegetation state. A neural net approach could be used for chlorophyll estimation as well, and in some cases it produced even better estimates than regression. However, in general it provided in this study somewhat lower quality of estimates than regression. This result could be explained by the fact that we used conventional net structures. Perhaps, special types of nets could be adjusted for this specific problem to obtain better results, and further research in this direction is required. Another promising approach to approximation of the first derivative curves could be the use of orthogonal polynomials, and application of image recognition techniques for estimation of chlorophyll content in vegetation. We got positive preliminary results of application of such approaches to estimation of chlorophyll concentration, and previously excellent results had been received in performing medical and engineering diagnosis.

REFERENCES

1. *Benedict H.M., Swidler R.* Nondestructive method for estimating chlorophyll content of leaves // Science. — 1961. — **133**. — P. 2015–2016.
2. *Hardwik K., Baker N.R.* In vivo measurement of chlorophyll content of leaves // New Phytol. — **72.** — P. 51–54, 1973.
3. *Carter G.A.* Ratios of leaf reflectance in narrow wavebands as indicators of plant stresses // Int. J. Rem. Sens.— 1994. — **15.** — P. 697–703.
4. *Yoder B.J., Pettigrew-Crosby R.E.* Predicting nitrogen and chlorophyll content and concentration from reflectance spectra (400–2500 nm) at leaf and canopy scales // Rem. Sens. Envir. — 1995. — **53**. — P. 199–211.
5. *Gitelson A.A., Merzlyak M.N.* Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll // J. Plant Physiol. — 1996. — **148.** — P. 494–500.
6. *Datt B.* A new reflectance index for remote sensing of chlorophyll content in higher plants: test using Eucalyptus leaves // J. Plant Physiol. — 1999. — **154.** — P. 30–36.
7. *Huete A.R.* A soil adjusted vegetation index (SAVI) // Rem. Sens. Envir. — 1988. — **25**. — P. 295–309.

8. *Qi J., Huete A.R.* Modified soil adjusted vegetation index (MSAVI) // Rem. Sens. Envir. — 1994. — **48**. — P. 119–126.

9. *Li Xia.* A two-axis adjusted vegetation index (TWVI) // Int. J. Rem. Sens. — 1994. — **14**. — P. 1447–1458.

10. *Collins W.* Remote sensing of crop type and maturity // Photogr. Eng. Rem. Sens. — 1978. — **44**. — P. 43–55.

11. *Horler D.N.H., Dokrey M., Barber J.* The red edge of plant leaf reflection // Int. J. Rem. Sens. — 1983. — № **4.** — P. 273–288.

12. *Miller J.R., Hare E.W., Wu J.* Quantitative characterization of the vegetation red edge reflectance. 1. An inverted-Gaussian reflectance model // Int. J. Rem. Sens. — 1990. — № 11. — P. 1755–1773.

13. *Joumar A.Yu.* Research of influence of parameters of vegetation canopy on the position of maximal declination of red edge // Research Earth from Space. — 1992. — № 4. — P. 15–20.

14. *Demetriodes-Shan T.H., Steven M.D., Clark J.A.* High resolution derivative spectra in remote Sensing // Rem. Sens. Envir. — 1993. — **33**. — P. 55–64.

15. *Kochubey S.M., Kobets N.I., Shadchina T.M.* The quantitative analysis of shape of spectral reflectance curves of plant leaves as a way for testing their status // Physiology and biochemistry of cultivated plants. — 1988. — **20.** — P. 535–539.

16. *Кочубей С.М., Кобец Н.И., Шадчина Т.М.* Спектральные свойства растений как базис для дистанционных методов диагностики. — Киев: Наук. думка, 1990. — 260 с.

17. *Kochubey S.M.* Comparative analysis of information power of multispectral imaging and high-resolution spectrometry in the remote sensing of vegetation cover // Space Sciences and Technology. — 1999. — **5**. — P. 41–48.

18. *Arnon D.I.* Copper enzymes in isolated chloroplasts // Polyphenoloxidase in Beta vulgaris Plant Physiol. — 1949. — **24.** — P. 1–15.

19. *Savitzky A., Golay M.G.E.* Smoothing and differentiation of data by simplified least squares Procedures // Analyt. Chem. — 1964. — **36.** — P. 1627–1639.

20. *Max J.* Methodes et techniques de traitement du signal et applications aux mesures physiques. — Paris: Masson, 1981. — 375 p.

21. *Дьяконов В., Круглов В.* Математический инструментарий Matlab. — Санкт-Петербург: Питер, 2001. — 280 с.

22. *Применение* искусственных нейронных сетей для прогнозирования длины волны в максимуме генерации лазерных красителей класса имитринов / В.И. Литвиненко, Ю.А. Бюргер, В.П. Кругленко, Ю.Ю. Стойлов // 8-я Всероссийская конф. по нейрокомпьютерам, НКП-2002, Москва, 21–22 марта 2002. — М., 2002. — С. 232–238.

———————————————

From the Editorial board: The article corresponds completely to submitted manuscript.